RELATION OF PREFERENCE PANEL AND TRAINED PANEL SCORES ON DRY WHOLE MILK [1]

LOIS A. SATHER AND L. D. CALVIN

Department of Food Science and Technology, Oregon State University, Corvallis

AND

A. TAMSMA

Dairy Products Laboratory, Eastern Utilization Research and Development Division, USDA
Washington, D. C.

SUMMARY

Dry whole milk samples manufactured by the EURDD Dairy Products Laboratory, Washington, D. C., were scored by an Oregon State University trained panel, EURDD Dairy Products Laboratory, Washington, D. C., trained panel, EURDD Engineering and Development Laboratory, Wyndmoor, Pennsylvania, trained panel, and by an OSU preference panel of from 100 to 160 student judges. After preliminary testing of 63 samples by the DPL and OSU panels, two series of 50 samples each for a total of 100 samples were tested by all four panels, except the Wyndmoor panel, which tested only 96 samples.

Two of the three trained panels used modifications of the American Dairy Science Association score-card for dry whole milk. The Oregon trained panel used a simple six-factor ballot. The student preference panel scored the samples on the nine-point hedonic scale.

All data were statistically analyzed for correlation and regression coefficients. Results indicated that total scores from the trained panels could be used as a basis for a scoring system which would predict consumer preference scores. The trained panels' scores for individual scoring factors, i.e., cooked, oxidized, stale, astringent flavors, did not satisfactorily predict preference scores, but could be used as an indication of possible flavor defects in dry whole milk samples.

Investigations on the manufacturing procedures and the storage stability of vacuum foam dried whole milk powders (2-4) have indicated a need for a study of the relationship between trained panel scores for flavor defects and consumer preference ratings. These investigations also have indicated that the principle flavor defects in dry whole milk, as scored by a trained panel, may be classified as oxidized, cooked, stale, and astringent. A previous study by Calvin and Sather (1) indicated that student preference panels could be used as an indication of consumer preference. Therefore, this study was undertaken to determine the correlation between trained panel scores for dry whole milks having oxidized, cooked, stale, and astringent flavor defects and student preference panel scores.

EXPERIMENTAL PROCEDURES

All dry whole milk samples were manufactured by the EURDD, Dairy Products Laboratory, Washington, D. C. The powders were prepared, packaged, and stored under varying conditions so as to induce flavor changes which would be representative of the range in quality which might be encountered under commercial manufacturing and distribution procedures. The variables introduced to obtain quality differences were pasteurization at 165, 180, or 195 F for 15 sec, moisture content 2 or 5%, oxygen in the interstitial gas .01 or 1%, storage temperatures 0, 40, or 80 F.

Samples from each lot were forwarded to the Department of Food Science and Technology, Oregon State University (OSU), Corvallis; to EURDD Engineering and Development Laboratory, Wyndmoor, Pennsylvania; and samples were retained by the Dairy Products Laboratory (DPL). Upon receipt, the samples were placed in 0 F storage until tested by

trained panels at each of the three respective laboratories and by a student preference panel at OSU.

All samples were reconstituted the day before they were to be tested. For small trained panel tests, 75 g of powder and 500 ml of distilled water (OSU) or spring water (DPL and Wyndmoor) were blended for approximately 60 sec in a Waring Blendor. For the larger student preference panel tests, the powders were reconstituted in a 2-gal electric butter churn, using 90 F distilled water and the same ratio of powder to water. The resulting fluid milk samples contained 12.6, ± 0.3% solids. The fluid milk samples were stored overnight in household refrigerators at approximately 45 F. For serving to the trained panels, the samples were heated in a hot water bath to 90 F, then cooled to 70 F. The preference panel samples were served at refrigerator temperature to the judges.

At OSU, the samples were served in coded 6-oz glasses to the judges, seated in individual testing booths, each containing a sink with water available and lighted with a yellow-orange light to mask any color differences between samples. Similar testing conditions were used by the DPL and Wyndmoor panels, except the samples were served in red plastic beakers.

Ten to fourteen judges served on each of the OSU, DPL, and Wyndmoor trained panels. These judges had been selected because of their acuity in detecting small differences in flavor in fluid milk samples and had participated in training sessions where samples were scored and then discussed. The OSU preference panel consisted of from 100 to 160 students. The only qualification for selection was that the student normally liked and consumed milk.

The scoring systems used by the DPL and Wyndmoor panels were modifications of the American Dairy Science Association (A.D.S.A.) score-card for dry whole milk. However, only the total scores and the scores for cooked, oxidized, stale, and astringent flavors for each samples were included in the analyses, as these were the principal criticisms indicated for the samples. On these scoring systems, the judge indicated the criticism or flavor defect as S (slight), D (definite), or P (pronounced), then gave a total score based on the A.D.S.A. scoring system. For the purpose of analyzing and comparing panel results, for cooked, oxidized, and stale flavors, the scores of 1, 2, 3 were assigned to the terms slight, definite, and pronounced, respectively. For astringent, the per cent of the panel indicating the presence of this factor was determined. The OSU trained

panel ballot used in the first series of tests asked the judge to score the intensity of cooked, oxidized, and stale flavors on a four-point scale from 0, none, to 4, pronounced. The OSU preference panel scored all samples on the nine-point hedonic scale on which 1 indicated dislike extremely, 5, neither like nor dislike, and 9, like extremely.

After preliminary testing on three consecutive weeks of three lots of a total of 35 samples by the OSU and DPL trained panels, the first series of 50 dry whole milk samples was scored by all three trained panels and the student preference panel. As a result of these tests, the OSU trained panel ballot was redesigned to include other flavors and total score as shown in Table 1 and used for all further tests. After

TABLE 1

OSU trained panel ballot no. 2

Flavor ballot

Sample No._____ Milk Name:_____

	Cooked	Oxidized	Stale	Other	Astringent
None					Yes
Slight					No
Definite					
Pronounced					
Total score	1 2 3 4 5 6 7 8 9 10				
	Low		High		

preliminary testing by the OSU and DPL trained panels of 28 samples, another series of 50 dry whole milk samples was tested by all four panels, except the Wyndmoor panel, which tested only 46 samples.

Mean scores were determined for each sample in each test and the data statistically analyzed, to determine the correlation coefficients between the trained panels' scores and between the trained panels' scores and the preference panel scores. Throughout the testing periods, all panels included fresh fluid milk samples as controls, but the scores were not included in the analyses, because these samples varied between panels.

RESULTS AND DISCUSSION

The correlation coefficients for the first three lots of 35 dry whole milk samples tested only by the OSU and DPL trained panels are given in Table 2.

These two trained panels were in good agreement on the scoring of cooked and oxidized flavors, as shown by the significant correlation

TABLE 2

Correlation coefficients
Dry whole milk
DPL and OSU panels

Between OSU and DPL trained panels	Cooked	Oxidized	Stale	Astringent
Lot 1, 10 samples	.82*	.95*	.67	.76*
Lot 2, 8 samples	.83*	.80*	−.68	−.15
Lot 3, 17 samples	.90*	.96*	.54	−.01

* Significant at the 1% level.

coefficients. There was no significant agreement between the two panels on the scoring of stale flavor and only on the first lot of ten samples when scoring astringent flavor.

For the first series of 50 dry whole milk samples judged by all four panels, the correlation coefficients between the trained panels' scores and the preference scores, the range of mean scores, and the number of samples actually scored as containing a given factor are given in Table 3.

from −.44, as shown in Table 3, for the highest individual factor to .63 and .82, as shown in Table 4, when all four factors were used.

As total score was the individual scoring factor which gave the highest correlation with preference scores, the OSU trained panel ballot was changed to include total score and another preliminary series of 27 samples tested by the OSU and DPL panels only. A correlation coefficient of 0.72 was obtained between the DPL and the OSU trained panels' total

TABLE 3

Correlation coefficients
Series 1, 50 dry whole milk samples

Factor and panel	No. of samples actually scored as containing factor	Range of mean scores	Corr. coeff. with preference scores
Preference panel scores	50	2.83- 6.35	
Cooked, Wynd.	46	0- 1.87	−.01
Cooked, DPL	46	0- 2.10	−.04
Cooked, OSU	49	0- 2.00	−.35
Oxidized, Wynd.	38	0- 1.10	−.53*
Oxidized, DPL	12	0- 1.00	−.44*
Oxidized, OSU	42	0- 0.90	−.01
Stale, Wynd.	49	0- 1.33	−.47*
Stale, DPL	50	0.20- 1.20	−.32
Stale, OSU	50	0.10- 1.70	−.44*
Astringent, Wynd.	8	0-12%	
Astringent, DPL	50	10-40%	.24
Astringent, OSU	35	0-60%	−.01
Total score, Wynd.	50	30.78-37.88	.65*
Total score, DPL	50	28.80-36.90	.65*

* Significant at the 1% level.

The only factor giving relatively high correlations with preference scores was total score, as scored by the Wyndmoor and DPL trained panels.

A multiple regression analysis was run on the OSU and DPL scores, using all four scoring factors to predict preference scores. The regression coefficients are given in Table 4.

The correlation coefficient between preference score and scoring factors was increased

TABLE 4

Regression coefficients
Series 1, 50 samples

	DPL	OSU
Cooked	− .47*	−1.01*
Oxidized	−2.16*	− .32
Stale	−1.27*	−1.61*
Astringent	.00	.00
Multiple correlation coefficient	.63*	.82*

* Significant at the 1% level.

scores on these 27 samples. Therefore, a second series of 50 samples was prepared and tested by all four panels, except Wyndmoor, which tested only 46 samples. The same samples were tested on the same day by each panel. The correlation coefficients, range of mean scores, and number of samples actually scored as containing a given factor for this second series of 50 samples are given in Table 5.

the correlation coefficients being so divergent.

To determine which scoring factors used by the trained panels had the greatest influence on preference score, regression analyses were made. The regression coefficients are given in Table 6.

Total score contributes by far the major portion of predictability when using all five scoring factors. The increase in the multiple correlation coefficient using five factors is less

TABLE 5

Correlation coefficients
Series 2, 50 dry whole milk samples

Factor and panel	No. of samples actually scored as containing factor	Range of mean scores	Corr. coeff. with preference scores
Preference panel scores	50	2.58- 6.03	
Cooked, Wynd.	45	0- 1.25	.42*
Cooked, DPL	37	0- 1.70	.43*
Cooked, OSU	50	0.18- 1.10	−.10
Oxidized, Wynd.	46	0.07- 1.57	−.76*
Oxidized, DPL	27	0- 1.80	−.74*
Oxidized, OSU	49	0- 1.36	−.65*
Stale, Wynd.	44	0- 0.83	−.57*
Stale, DPL	48	0- 1.10	−.19
Stale, OSU	50	0.22- 1.50	−.67*
Astringent, Wynd.	3	0- 8%	
Astringent, DPL	50	10-50%	.40*
Astringent, OSU	44	0-64%	−.78*
Total score, Wynd.	46	29.36-38.50	.86*
Total score, DPL	50	24.70-37.20	.86*
Total score, OSU	50	2.30- 7.30	.88*

* Significant at the 1% level.

When comparing the trained panels' scores with preference scores, oxidized flavor and total score consistently have the highest correlation coefficients.

Correlation coefficients from Series 2 of 0.86 or better (Table 5) between trained panels' total scores and preferences scores would indicate a relatively high ability to predict. Correlation coefficients of 0.65 from Series 1 (Table 3) do not indicate a very high ability to predict. Series 1 samples were more limited in range of quality than the samples in Series 2. This is probably the primary difference for

than .05 as compared to using total score alone, as shown in Table 5.

As the three trained panels were using different total scoring systems, the equivalent preference scores were computed using the following formula obtained from the regression equation relating total score to preference score:

$$Y = \text{preference score}$$
$$X = \text{total score}$$

Wyndmoor $Y = -7.78 + .353\,X$
DPL $\quad\quad Y = -5.00 + .290\,X$
OSU $\quad\quad Y = 1.83 + .566\,X$

TABLE 6

Regression coefficients
Series 2, 50 samples

	DPL	Wyndmoor	OSU
Cooked	.03	−.24	−1.23*
Oxidized	−.37	−.59	− .30
Stale	−.25	.04	.48
Astringent	.00	.00	.00
Total score	.23*	.29*	.60*
Multiple correlation coefficient	.87*	.88*	.93*

* Significant at the 1% level.

[4]

In Table 7 are given the total scores for each panel for preference scores from three to seven on the hedonic scale.

Thus, any of the three total scoring systems used by the three trained panels may be used in predicting preference scores.

CONCLUSIONS

Total score, the over-all evaluation, was found to be the best quantity for flavor evaluation of milk by trained panels for predicting consumer preferences.

The different scoring systems used by the three trained panels did not affect the validity of using total scores to predict consumer preference.

A simple scoring system such as used by the Oregon trained panel appears adequate and probably could be used as a basis for standardizing flavor evaluation of milk by trained panels.

Standardization of flavor evaluation procedures is needed for proper interpretation of flavor results obtained by different panels in different locations and periods.

Dry milk samples are suitable to compare the results of different flavor panels.

REFERENCES

(1) CALVIN, LYLE D., AND SATHER, LOIS A. A Comparison of Student Preference Panels with a Household Consumer Panel. Food Technol., 8: 469. 1959.

(2) TAMSMA, A., MUCHA, T. J., AND PALLANSCH, M. J. Factors Related to the Flavor Stability During Storage of Foam-Dried Milk. II. Effect of Heating Milk Prior to Drying. J. Dairy Sci., 45: 1435. 1962.

(3) TAMSMA, A., MUCHA, T. J., AND PALLANSCH, M. J. Factors Related to the Flavor Stability During Storage of Foam-Dried Whole Milk. III. Effect of Antioxidants. J. Dairy Sci., 46: 114. 1963.

(4) TAMSMA, A., PALLANSCH, M. J., MUCHA, T. J., AND PATTERSON, W. I. Factors Relating to the Flavor Stability of Foam-Dried Whole Milk. I. Effect of Oxygen Level. J. Dairy Sci., 44: 1644. 1961.

TABLE 7
Total scores for equivalent preference scores

Preference score	3	4	5	6	7
Total score					
Wyndmoor	30.59	33.42	36.25	39.08	41.91
DPL	27.59	31.04	34.49	37.94	41.39
OSU	2.08	3.85	5.62	6.39	9.16